

# VIDEO GENERATING SYSTEM AND CUSTOM VIDEO GENERATING METHOD

**Publication number:** JP2002084488 (A)

**Also published as:**

**Publication date:** 2002-03-22

US6807361 (B1)

**Inventor(s):** ANDREAS GARGENSHORN; JOHN J DOHERTY; WILCOX LYNN D; JOHN S BORECKY; CHIU PATRICK; JONATHAN T FOOTE

**Applicant(s):** FUJI XEROX CO LTD

**Classification:**

- international: H04N5/91; G11B27/034; G11B27/28; G11B27/34; H04N5/91; G11B27/031; G11B27/28; G11B27/34; (IPC1-7): H04N5/91

- European: G11B27/034; G11B27/28; G11B27/34

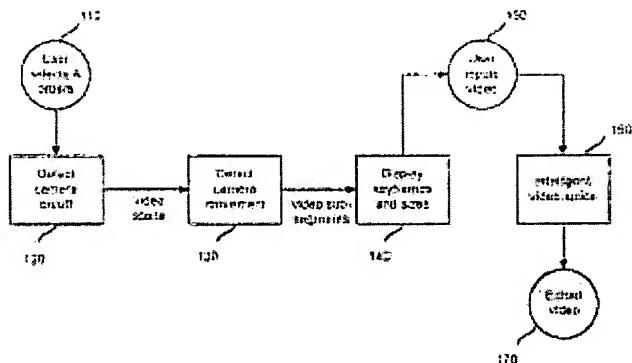
**Application number:** JP20010208903 20010710

**Priority number(s):** US20000618533 20000718

---

## Abstract of JP 2002084488 (A)

**PROBLEM TO BE SOLVED:** To provide an interactive video generating system that analyzes and synthesizes video data without the need for manual selection of start and end points by each clip. **SOLUTION:** The system of this invention analyzes video to decide a camera motion and a camera speed. A video clip suitably to be included in a final video is identified by using hierarchical edit rules and a key frame for each suitable clip is displayed on a story board for users. Then key frames are sequenced in a desirable sequence. By connecting selected clips, a final output video is generated.



---

Data supplied from the [esp@cenet](mailto:esp@cenet) database — Worldwide

---

(19)日本特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-84488

(P2002-84488A)

(43)公開日 平成14年3月22日 (2002.3.22)

(51)Int.Cl.<sup>7</sup>

識別記号

H 0 4 N 5/91

F I

データコード\*(参考)

H 0 4 N 5/91

N 5 C 0 5 3

審査請求 未請求 請求項の数32 O.L (全 10 頁)

(21)出願番号 特願2001-208903(P2001-208903)  
(22)出願日 平成13年7月10日(2001.7.10)  
(31)優先権主張番号 6 1 8 5 3 3  
(32)優先日 平成12年7月18日(2000.7.18)  
(33)優先権主張国 米国(US)

(71)出願人 000005496  
富士ゼロックス株式会社  
東京都港区赤坂二丁目17番22号  
(72)発明者 アンドレアス ガーゲンショーン  
アメリカ合衆国 94025 カリフォルニア  
州 メンロ パーク ウェイヴァリー ス  
トリート 210 アパートメント 4  
(72)発明者 ジュン ジエイ、ドゥハーティ  
アメリカ合衆国 94116 カリフォルニア  
州 サンフランシスコ ヴィセンテ スト  
リート 1524  
(74)代理人 100079049  
弁理士 中島 淳 (外1名)

最終頁に続く

(54)【発明の名称】 ビデオ生成システム及びカスタムビデオ生成方法

(57)【要約】

【課題】各クリップ毎に開始ポイントと終了ポイントとの手作業での選択を必要としないビデオ解析し及び合成する対話式ビデオ生成システムが必要である。

【解決手段】本システムは、ビデオを解析してカメラモーションとスピードを決定する。最終ビデオに含まれるのが適するビデオのクリップは、複数の階層的編集ルールを使用して識別され、各好適なクリップ毎のキーフレームは、ユーザに対してストーリボードにディスプレイされる。次に、キーフレームは、望ましいシーケンスで順序付けられる。最終出力ビデオは、選択されたクリップを繋げることによって生成される。

【特許請求の範囲】

【請求項1】 ビデオカメラからの記録されたビデオを解析して最終出力ビデオを製作するビデオ生成システムであって、前記システムは、

前記記録されたビデオの夫々のシーンをクリップにセグメント化するカメラモーション検出器を備え、前記クリップが夫々のシーンで検出されたカメラモーションに従ってクラス分けし、

夫々のクリップを表すキーフレームに隣接するビデオフレームが共にスプライスされると最終出力ビデオを生成するか否かを決定するために複数の編集ルールを適用するビデオスプライサを、備える、ビデオ生成システム。

【請求項2】 デジタイザーを更に備え、前記デジタイザーは、前記記録されたビデオをデジタル化し、且つ前記記録されたビデオをカメラのオンオフ移行に対応するシーンにセグメント化する、請求項1に記載のシステム。

【請求項3】 ユーザインターフェースを更に備え、前記ユーザインターフェースは、夫々の選択されたキーフレームをディスプレイし、且つユーザが一連のキーフレームを希望する順序に変更することを可能とする、請求項1に記載のシステム。

【請求項4】 前記ユーザインターフェースは、キーフレームインターフェースとストーリボードインターフェースを備える、請求項3に記載のシステム。

【請求項5】 前記キーフレームインターフェースにディスプレイされる夫々のキーフレームのサイズは、前記対応するキーフレームによって表される夫々のクリップの期間に比例する、請求項4に記載のシステム。

【請求項6】 ユーザは、前記キーフレームインターフェースにディスプレイされる夫々のキーフレームを希望のタイムシケンスでストーリボード上に配置出来る、請求項4に記載のシステム。

【請求項7】 前記ユーザインターフェースは、ダイナミックユーザインターフェースであり、前記ダイナミックユーザインターフェースは、前記記録されたビデオが再生されている間に、ユーザが前記最終出力ビデオから見られているクリップをオミットすることを可能とする、請求項3に記載のシステム。

【請求項8】 前記ユーザインターフェースは、ユーザが見ているクリップを前記最終出力ビデオに含ませる事を可能とする請求項3に記載のシステム。

【請求項9】 キーフレームセレクタを更に備え、前記キーフレームセレクタは、ユーザが夫々のクリップを表す少なくとも一つのキーフレームを選択することを可能とする、請求項1に記載のシステム。

【請求項10】 単一のキーフレームが、スタティックシーンを有するスチルクラスとしてクラス分けされたクリップに対して、選択され且つディスプレイされる、請求項9に記載のシステム。

【請求項11】 複数キーフレームが、オブジェクトモーションを中心に有するクリップに対して選択且つディスプレイされる、請求項9に記載のシステム。

【請求項12】 複数キーフレームが、階層的集塊クラスタリングを使用して夫々のクリップを同質領域にセグメント化し、且つキーフレームを夫々の領域から選択することによって、選択される、請求項11に記載のシステム。

【請求項13】 前記ビデオスプライサは、前記複数の編集ルールを適用するための制約条件満足システムを使用する、請求項1に記載のシステム。

【請求項14】 エルゴード隠れマルコフモデルは、前記カメラモーション検出器によってシーン中に検出されるカメラモーションクラスに基づいて、夫々のシーンをクリップにセグメント化するために使用される、請求項1に記載のシステム。

【請求項15】 夫々のクリップは、スチルクラス、パンクラス、チルトクラス、ズームクラス、及びガーベージクラスより成るカメラモーションに従って、クラス分けされる、請求項1に記載のシステム。

【請求項16】 前記編集ルールは、  
3秒に実質的に等しい最小長よりも短い長さを有する夫々のクリップを破棄することと、  
10秒に実質的に等しい最大長を越える長さを有する夫々のクリップをトリミングすることと、  
最終出力ビデオに含まれると共に、同じシーンから選択される二つのクリップであって、前記二つのクリップが3秒未満だけ分離されている場合、これらの二つのクリップ間を切断することを回避するために、前記二つのクリップを併合することと、  
高速で非線形カメラモーションを有するクリップとガーベージクラスとしてクラス分けされたクリップを破棄することと、

ショットが前記最大長を超える場合このショットの終わり近くのサブクリップを選択することと、  
スチルクリップがズーム、パン及びチルトの何れかの端部に存在する場合、5秒未満の期間を有する前記ズーム、パン及びチルトよりなるクリップを破棄することと、

ショットが所定の明るさ閾値よりも高い最小明るさを有する場合そのショットを選択すること、  
を備える、請求項1に記載のシステム。

【請求項17】 前記ビデオスプライサは、前記編集ルールを適用して自動的に夫々のクリップに対してビデオインポイントとビデオアウトポイントを決定する、請求項1に記載のシステム。

【請求項18】 ビデオカメラの記録されたビデオからカスタムビデオを生成する方法であって、前記方法は、前記記録されたビデオの夫々のシーン中にカメラ移動を検出するステップと

検出されたカメラ移動のクラスに基づいて夫々のシーンをクリップにセグメント化するステップと、夫々のクリップを表す選択されたキーフレームをユーザインターフェース上にディスプレイするステップであって、夫々のキーフレームの前記選択は、検出されたカメラ移動のクラスに基づくステップと、編集ルールを前記ユーザインターフェースにディスプレイされたキーフレームへ適用して夫々のキーフレーム周りのビデオのセクションを選択し共にスプライスして最終出力ビデオを生成するステップと、を備えるカスタムビデオ生成方法。

【請求項19】 前記記録されたビデオをデジタル化し且つ前記デジタル化され記録されたビデオをシーンにセグメント化するステップを更に備える、請求項18に記載の方法。

【請求項20】 夫々のシーンは、カメラオン移行とカメラオフ移行との間のデジタル化され記録されたビデオによって画定される、請求項19に記載の方法。

【請求項21】 エルゴード隠れマルコフモデルは、カメラ移動検出ステップによって前記シーン中で検出されたカメラ移動クラスに基づいて、夫々のシーンをクリップにセグメント化する、請求項18に記載の方法。

【請求項22】 前記カメラ移動クラスは、スチルクラス、パンクラス、チルトクラス、ズームクラス、及びガーベージクラスより成る、請求項21に記載の方法。

【請求項23】 単一キーフレームは、スタティックシーンを有するスチルクラスとしてクラス分けされたカメラモーションよりなるクリップに対して選択される、請求項18に記載の方法。

【請求項24】 複数のキーフレームは、オブジェクトモーションを有する、パンクラス、チルトクラス、ズームクラス、及びスチルクラスとして夫々クラス分けされたカメラモーションよりなるクリップに対して選択される、請求項18に記載の方法。

【請求項25】 前記複数のキーフレームは、階層的集塊クラスタリングを使用して夫々のクリップを同質領域へ分割すること及びキーフレームを夫々の領域から選択することによって選択される、請求項24に記載の方法。

【請求項26】 前記ユーザインターフェースにディスプレイされる夫々の選択されたキーフレームのサイズは、それぞれの選択されたキーフレームによって表示されるクリップの期間に比例する、請求項18に記載の方法。

【請求項27】 前記ユーザインターフェースは、キーフレームインターフェースとストーリボードインターフェースとを備え、ユーザは、前記キーフレームインターフェース上にディスプレイされる夫々の選択されたキーフレームを希望のシーケンスで前記キーフレームボード上に配する事が出来る、請求項26に記載の方法。

【請求項28】 前記ユーザインターフェースは、第1のディスプレイインターフェースと第2のディスプレイインターフェースとを備える、請求項18に記載の方法。

【請求項29】 前記スタティックユーザインターフェースは、選択されたキーフレームをディスプレイし、且つ、前記記録されたビデオが再生されている間に、前記ダイナミックユーザインターフェースは、ユーザが前記最終出力ビデオからオミットされるクリップをダイナミックに選択する、請求項28に記載の方法。

【請求項30】 前記編集ルールを適用するステップは、更に、

3秒に実質的に等しい最小長よりも短い長さを有する夫々のクリップを破棄するステップと、

10秒に実質的に等しい最大長を越える長さを有する夫々のクリップをトリミングするステップと、

最終出力ビデオに含まれると共に、同じショットから選択される二つのクリップであって、前記二つのクリップが3秒未満だけ分離されている場合、これらの二つのクリップ間を切断することを回避するために、前記二つのクリップを併合するステップと、

高速で非線形カメラモーションを有するクリップとガーベージクラスとしてクラス分けされたクリップを破棄するステップと、

ショットが前記最大長を超える場合このショットの終わり近くのサブクリップを選択するステップと、

スチルクリップがズーム、パン及びチルトの何れかの端部に存在する場合、5秒未満の期間を有する前記ズーム、パン及びチルトよりなるクリップを破棄するステップと、

前記ショットが所定の明るさ閾値よりも高い最小明るさを有する場合そのショットを選択するステップと、を備える、請求項18に記載の方法。

【請求項31】 前記編集ルールを適用するステップは、更に、夫々のクリップに対してビデオインポイントとビデオアウトポイントを自動的に決定するステップを備える、請求項18に記載の方法。

【請求項32】 前記セグメント化ステップは、別法として、ビデオ品質ルールに基づいて夫々のシーンをセグメント化する、請求項18に記載の方法。

#### 【発明の詳細な説明】

##### 【0001】

【発明の属する技術分野】本発明は、ビデオカメラからの記録されたビデオを解析するためのビデオ生成システム及びビデオカメラの記録されたビデオからカスタムビデオを生成するための方法に関する。

##### 【0002】

【従来の技術】ビデオカメラは、家庭での使用及びオフィスでの使用の両方のために益々普及している。ビデオは、バケーション、結婚式及び卒業式のようなファミリ

イベントを記録するための使用される。オフィスの場合、ビデオカメラは、プレゼンテーションとユーザ研修を記録するために使用され、且つしばしば人々、場所及びミーティングを記録するためにビジネストリップに持って行かれる。更に、ディジタルビデオ(DV)、ディジタルカムコーダ(camcorder)フォーマットの使用は、高品質ディジタルビデオ、安価なビデオ捕獲カード及びハードディスクの組合せ及びワールドワイドウェブ(WWW)に対するビデオコンテンツ生成における公共の利益に起因して発展している。

【0003】しかしながら、問題は、ビデオが記録された後そのビデオを利用する事が困難であることである。人々が、記録されたビデオを一度又は二度見るかもしれないが、それは、一般に”袋の中に”しまったままとなる。理由は、興味ある部分は、しばしばより長い、より短い興味ある部分と混合されるからである。例えば、ビデオの品質低下がしばしば急激なカメラ移動や見る人の興味を長く維持できないビデオ部分から生じる。更に、特に、望ましい部分がビデオの他の部分よりも短い場合、又はビデオの異なる部分が異なる人々によって興味が持たれる場合、早送り及び早戻しのみを使用して興味ある部分を検出することが困難である。

【0004】ビデオ編集ツール(例えば、Adobe PremierやIn-Sync's Speed Razer)があるが、これらのツールは、アニメータが使用するのが非常に難しい。AppleのiMovie、Ulead SystemのVideoStudioやJavuNetworkのようなより単純なビデオ編集ツールは、ビデオシーン(take)の選択及びそれらのストーリーボードへの配置を容易にする。しかしながら、これらのシステムの何れもが、知的編集援助を行なうためのビデオ解析を実行しない。特に、これらのシステムの大部分が、シーンをトリムするのに必要なインポイントとアウトポイントを選択するための援助を提供せず、従って、ユーザは、手作業でフレーム毎にビデオを調べそして手作業でシーンから望ましくない部分をトリムしなければならない。この作業は、より単純な編集ツールを使用しても困難である。

#### 【0005】

【発明が解決しようとする課題】従って、各クリップ毎に開始ポイントと終了ポイントとの手作業での選択を必要としないビデオ解析し及び合成する対話式ビデオ生成システムが必要である。ビデオを解析して適切なクリップを識別し且つ一セットの階層的編集ルールを適用して自動的に各クリップ毎に開始ポイントと終了ポイントを決定する対話式ビデオ生成システムが必要である。

#### 【0006】

【課題を解決するための手段】概略的には、本発明は、ビデオカメラの記録されたビデオからビデオを合成するための対話式ビデオ生成システムを提供し、それは各ビ

デオクリップ毎に開始及び終了ポイントを自動的に決定する。このシステムは、ビデオを解析してカメラモーションとスピードを決定する。仕上げビデオに含まれるのが適するビデオのセグメント、即ちクリップは、複数の階層的編集ルールを使用して識別され、各好適なクリップ毎のキーフレームは、ユーザに対してストーリーボードにディスプレイされる。次に、キーフレームは、望ましいシーケンスで順序付けられる。最終出力ビデオは、選択されたクリップを繋げることによって生成される。

【0007】具体的には、本発明の第1の態様は、ビデオカメラからの記録されたビデオを解析して最終出力ビデオを製作するビデオ生成システムであって、前記システムは、前記記録されたビデオの夫々のシーンをクリップにセグメント化するカメラモーション検出器を備え、前記クリップが夫々のシーンで検出されたカメラモーションに従ってクラス分けし、夫々のクリップを表すキーフレームに隣接するビデオフレームが共にスプライスされて最終出力ビデオを生成するか否かを決定するために複数の編集ルールを適用するビデオスライサを、備えるビデオ生成システムである。

【0008】第2の態様は、上記第1の態様において、デジタイザーを更に備え、前記デジタイザーは、前記記録されたビデオをデジタル化し、且つ前記記録されたビデオをカメラのオンオフ移行に対応するシーンにセグメント化する。

【0009】第3の態様は、第1の態様において、ユーザインターフェースを更に備え、前記ユーザインターフェースは、夫々の選択されたキーフレームをディスプレイし、且つユーザが一連のキーフレームを希望する順序に変更することを可能とする。

【0010】第4の態様は、第3の態様において、前記ユーザインターフェースは、キーフレームインターフェースとストーリーボードインターフェースを備える。

【0011】第5の態様は、第4の態様において、前記キーフレームインターフェースにディスプレイされる夫々のキーフレームのサイズは、前記対応するキーフレームによって表される夫々のクリップの期間に比例する。

【0012】第6の態様は、第4の態様において、ユーザは、前記キーフレームインターフェースにディスプレイされる夫々のキーフレームを希望のタイムシーケンスでストーリーボード上に配置出来る。

【0013】第7の態様は、第3の態様において、前記ユーザインターフェースは、ダイナミックユーザインターフェースであり、前記ダイナミックユーザインターフェースは、前記記録されたビデオが再生されている間に、ユーザが前記最終出力ビデオから見られているクリップをオミットすることを可能とする。

【0014】第8の態様は、第7の態様において、前記ダイナミックユーザインターフェースは、スタティックユーザインターフェースと共に使用される。

【0015】第9の態様は、第3の態様において、前記ユーザインターフェースは、ユーザが見ているクリップを前記最終出力ビデオに含ませる事を可能とする。

【0016】第10の態様は、第1の態様において、キーフレームセレクタを更に備え、前記キーフレームセレクタは、ユーザが夫々のクリップを表す少なくとも一つのキーフレームを選択することを可能とする。

【0017】第11の態様は、第10の態様において、单一のキーフレームが、スタティックシーンを有するスチルクラスとしてクラス分けされたクリップに対して、選択され且つディスプレイされる。

【0018】第12の態様は、第10の態様において、複数キーフレームが、オブジェクトモーションを中心に有するクリップに対して選択且つディスプレイされる。

【0019】第13の態様は、第12の態様において、複数キーフレームが、階層的集塊クラスタリングを使用して夫々のクリップを同質領域にセグメント化し、且つキーフレームを夫々の領域から選択することによって、選択される。

【0020】第14の態様は、第1の態様において、前記ビデオスプライサは、前記複数の編集ルールを適用するための制約条件満足システムを使用する。

【0021】第15の態様は、第1の態様において、エルゴード隠れマルコフモデルは、前記カメラモーション検出器によってシーン中に検出されるカメラモーションクラスに基づいて、夫々のシーンをクリップにセグメント化するために使用される。

【0022】第16の態様は、第1の態様において、夫々のクリップは、スチルクラス、パンクラス、チルトクラス、ズームクラス、及びガーベージクラスより成るカメラモーションに従って、クラス分けされる。

【0023】第17の態様は、第1の態様において、前記編集ルールは、3秒に実質的に等しい最小長よりも短い長さを有する夫々のクリップを破棄することと、10秒に実質的に等しい最大長を越える長さを有する夫々のクリップをトリミングすることと、最終出力ビデオに含まれると共に、同じシーンから選択される二つのクリップであって、前記二つのクリップが3秒未満だけ分離されている場合、これらの二つのクリップ間を切断することを回避するために、前記二つのクリップを併合することと、高速で非線形カメラモーションを有するクリップとガーベージクラスとしてクラス分けされたクリップを破棄することと、ショットが前記最大長を超える場合このショットの終わり近くのサブクリップを選択することと、スチルクリップがズーム、パン及びチルトの何れかの端部に存在する場合、5秒未満の期間を有する前記ズーム、パン及びチルトよりなるクリップを破棄することと、ショットが所定の明るさ閾値よりも高い最小明るさを有する場合そのショットを選択すること、を備える。

【0024】第18の態様は、第17の態様において

前記所定の明るさ閾値は、実質的に30%の明るさに等しい。

【0025】第19の態様は、第1の態様において、前記ビデオスプライサは、前記編集ルールを適用して自動的に夫々のクリップに対してビデオインポイントとビデオアウトポイントを決定する。

【0026】第20の態様は、ビデオカメラの記録されたビデオからカスタムビデオを生成する方法であった、前記方法は、前記記録されたビデオの夫々のシーン中にカメラ移動を検出するステップと、検出されたカメラ移動のクラスに基づいて夫々のシーンをクリップにセグメント化するステップと、夫々のクリップを表す選択されたキーフレームをユーザインターフェース上にディスプレイするステップであって、夫々のキーフレームの前記選択は、検出されたカメラ移動のクラスに基づくステップと、編集ルールを前記ユーザインターフェースにディスプレイされたキーフレームへ適用して夫々のキーフレーム周りのビデオのセクションを選択し共にスライスして最終出力ビデオを生成するステップと、を備えるカスタムビデオ生成方法。

【0027】第21の態様は、第20の態様において、前記記録されたビデオをデジタル化し且つ前記ディジタル化され記録されたビデオをシーンにセグメント化するステップを更に備える。

【0028】第22の態様は、第21の態様において、夫々のシーンは、カメラオン移行とカメラオフ移行との間のディジタル化され記録されたビデオによって画定される。

【0029】第23の態様は、第20の態様において、エルゴード隠れマルコフモデルは、カメラ移動検出ステップによって前記シーン中で検出されたカメラ移動クラスに基づいて、夫々のシーンをクリップにセグメント化する。

【0030】第24の態様は、第23の態様において、前記カメラ移動クラスは、スチルクラス、パンクラス、チルトクラス、ズームクラス、及びガーベージクラスより成る。

【0031】第25の態様は、第20の態様において、单一キーフレームは、スタティックシーンを有するスチルクラスとしてクラス分けされたカメラモーションによるクリップに対して選択される。

【0032】第26の態様は、第20の態様において、複数のキーフレームは、オブジェクトモーションを有する、パンクラス、チルトクラス、ズームクラス、及びスチルクラスとして夫々クラス分けされたカメラモーションによるクリップに対して選択される。

【0033】第27の態様は、第26の態様において、前記複数のキーフレームは、階層的集塊クラスタリングを使用して夫々のクリップを同質領域へ分割すること及びキーフレームを夫々の領域から選択することによって

選択される。

【0034】第28の態様は、第20の態様において、前記ユーザインターフェースにディスプレイされる夫々の選択されたキーフレームのサイズは、それぞれの選択されたキーフレームによって表示されるクリップの期間に比例する。

【0035】第29の態様は、第28の態様において、前記ユーザインターフェースは、キーフレームインターフェースとストーリボードインターフェースとを備え、ユーザは、前記キーフレームインターフェース上にディスプレイされる夫々の選択されたキーフレームを希望のシーケンスで前記キーフレームボード上に配する事が出来る。

【0036】第30の態様は、第20の態様において、前記ユーザインターフェースは、第1のディスプレイインターフェースと第2のディスプレイインターフェースとを備える。

【0037】第31の態様は、第30の態様において、前記スタティックユーザインターフェースは、選択されたキーフレームをディスプレイし、且つ、前記記録されたビデオが再生されている間に、前記ダイナミックユーザインターフェースは、ユーザが前記最終出力ビデオからオミットされるクリップをダイナミックに選択する。

【0038】第32の態様は、第20の態様において、前記編集ルールを適用するステップは、更に、3秒に実質的に等しい最小長よりも短い長さを有する夫々のクリップを破棄するステップと、10秒に実質的に等しい最大長を越える長さを有する夫々のクリップをトリミングするステップと、最終出力ビデオに含まれると共に、同じショットから選択される二つのクリップであって、前記二つのクリップが3秒未満だけ分離されている場合、これらの二つのクリップ間を切断することを回避するために、前記二つのクリップを併合するステップと、高速で非線形カメラモーションを有するクリップとガーベージクラスとしてクラス分けされたクリップを破棄するステップと、ショットが前記最大長を超える場合このショットの終わり近くのサブクリップを選択するステップと、スチルクリップがズーム、パン及びチルトの何れかの端部に存在する場合、5秒未満の期間を有する前記ズーム、パン及びチルトよりなるクリップを破棄するステップと、前記ショットが所定の明るさ閾値よりも高い最小明るさを有する場合そのショットを選択するステップと、を備える。

【0039】第33の態様は、第20の態様において、前記編集ルールを適用するステップは、更に、夫々のクリップに対してビデオインポイントとビデオアウトポイントを自動的に決定するステップを備える。

【0040】第34の態様は、第20の態様において、前記セグメント化ステップは、別法として、ビデオ品質ルールに基づいて夫々のシーンをヤグメント化する。

【0041】第35の態様は、第34の態様において、夫々のシーンから最も高いビデオ品質を有するクリップを表すキーフレームが選択される。

【0042】本発明のこれらと他の目的、及び利点は、本発明の好適な実施の形態が添付図面と共に説明される以下の詳細な説明から一層明瞭になる。

【0043】

【発明の実施の形態】図1は、本発明による対話式ビデオ生成システムのブロック図である。ビデオカメラからの記録ビデオは、ブロック110に入力され、引き続いてブロック120でデジタル化され且つシーンにセグメント化される。ここで、シーン(take)は、カメラのオン移行とオフ移行との間に記録されたビデオである。カメラオンとオフの移行についての情報は、一般的には、アナログビデオカメラとは対照的に、デジタルビデオカメラによって提供されることが好適である。理由は、これはユーザの意図により近づくからである。アナログビデオカメラの場合、カメラのオンとオフの移行は、一般的に、標準ショット検出アルゴリズムによって記録ビデオから推定されなければならない。

【0044】各シーンは、ブロック130で解析され、そして以下で説明されるモーション検出アルゴリズムによって夫々のシーンで検出されたカメラ移動に従って複数のセグメントにセグメント化される。このモーション検出アルゴリズムによって検出可能なビデオのクラスは、スチル、パン、チルト、ズーム及びガーベージである。スチルは、2つのサブクラスに分解され、その内の方は、カメラが静止しているがオブジェクトのモーションがあり、他方は、オブジェクトモーションはない。パンはあらゆる連続する水平方向カメラ移動である。クラス分けに対して、パンは、2つのサブクラス、即ち、左パンと右パン、に分解される。パンと同様に、チルトも上方チルトと下方チルトの2つのサブクラスに分解される。ズームは、カメラ画像のスケール変化であり、ズームインとズームアウトの2つのサブクラスがある。ズームインでは、より大きな画像が得られ、ズームアウトでは、より小さな画像が得られる。パン、チルト及びズームでは、カメラの移動は比較的スローである。全ての高速又は非線形カメラ移動は、ガーベージクラスに入る。カメラ(又はオブジェクト)のタイプの知識は、編集ルールの実施とキーフレームの選択のために必要とされる。例えば、高速や誤ったカメラモーションを含むビデオクリップは、ビデオ品質に対応し、それらは削除されるべきである。パン、チルト及びズームでのゆっくりとしたカメラモーションが識別されるなければならない。理由は、複数のキーフレームがビデオのコンテンツを適切にディスプレイするために必要とされるからである。

【0045】隠れマルコフモデル(HMM)は、ビデオをカメラモーションのタイプに基づくクラスに対応する

領域にセグメント化するために使用される。エルゴードHMMが使用される。理由は、記録されたビデオがどのようにショットされたかについての前提が無いからである。図2は、HMMのトップレベル構造を示す。初期のヌル状態210から、スチルクラス220、パンクラス230、チルトクラス240、ズームクラス250、又はガーベージクラス260に移行する。ガーベージクラス260は、示されている単一状態によってモデル化される。これらの状態での自己移行は、そのクラスの期間をモデル化する。全ての他の状態は、サブHMMでモデル化されたサブクラスを有するクラスに対応する。また、パンのサブHMMの例が図3に示される。図3において、初期ヌル状態270からサブクラス状態であるパン右クラス280とパン左クラスの一方に移行し、これらの状態における自己移行が期間をモデル化する。

【0046】HMMの特徴ベクトルは、画像の複数のサブブロックに亘って演算されるモーションベクトルである。これらのベクトルは、ムービングピクチャーエクスパートグループ (Moving Picture Experts Group) (MPEG) 符号化から或いは他のモーション捕獲技術からのモーションベクトルである。HMMの特徴確率分布は、完全共分散ガウス分布である。完全共分散は、モーションのクラス分けの重要な要素である、モーションベクトル同士間の相関をモデル化するために必要である。パラメータは、クラスに従ってラベル付けされたトレーニングビデオから学習される。ガウス確率は、そのクラスのトレーニングデータの平均及び共分散マトリックスで初期化される。次に、これらのパラメータ及び遷移確率は、トレーニング手順の間にボームウェルチ (BaumWeich) アルゴリズムを使用して再予測される。

【0047】HMMがトレーニングされると、そのビデオの各シーンは、HMMを使用してクラスにセグメント化される。標準ビテルビアルゴリズムは、最も起こり得るシーケンスの状態、従って、各タイム毎にそのビデオに対する最も可能なクラスを決定するために使用される。次に、これらのクラスは、ビデオのクリップを定義し、そこでは、一つのクリップは、連続セットの連続フレーム (一つのフレームは、記録されたビデオから取られた单一のスチル画像である) である。また、他の特徴セットがHMMに対する特徴ベクトルとして使用されてもよい。例えば、2つの特徴は、各画像の9つのサブブロックのモーションベクトルから演算され得る (即ち、一方の特徴は、全て9つのモーションベクトルの大きさの合計であり、他方の特徴は、モーションベクトルの合計の大きさである)。この方法は、パン、チルト、及びズームの検出において合理的性能を生成するが、それは、オブジェクトモーションの存在下で低下する。

【0048】次に、キーフレームは、検出されたカメラ移動のタイプに基づいて クリップからブロック140

で選択される。キーフレームは、各クリップが一つ又はそれより多くのフレームによって表される、最終出力ビデオに対して選択された特定のビデオクリップを表すのに適するスチルフレームである。一つ又は複数のキーフレームは、スチル、パン、チルト又はズームとして先にクラス分けされたクリップから選択される。静止画像を有するスチル領域に対して、單一キーフレームが選択される。この場合、ショットの終り近くのキーフレームは、典型的なショットがその終了に向かって最も興味あることを指定する編集ルールに基づいて選択される。

【0049】パン、チルト、又はズームに対して、及びオブジェクトモーションを有するスチルに対して、複数のキーフレームは、ユーザがショットのコンテンツをよりよく制御出来るように選択される。複数のキーフレームは、階層的集塊クラスタリングを使用してその領域を等価領域にセグメント化し、次に各領域毎に一つのキーフレーム選択することによって選択される。このクラスタリングは、クリップ中の各フレームがそれ自体のクラスタであるように初期化される。隣合うフレーム間の距離は、ヒストグラム差に基づいて演算される。最も近接する2つの隣接フレームは、併合距離が閾値を超えるまで、反復的に併合される。得られたクラスタの何れかが予め設定された最小長さよりも短い場合、集塊クラスタリングは、そのクラスタに対して続けられる。最小長さは、ビデオクリップがあまりにも短すぎない事を指定する編集ルールから決定される。ビデオクリップに対する最小長さは、一般的には3秒であり、且つクラスタリングは、一般的には最小長さよりも短いクリップには実行されない。次に、キーフレームは、そのクリップの終り近くで、再び、得られたクラスタの各々に対して選択される。

【0050】選択されると、キーフレームは、図4に示されるように、キーフレーム中又はユーザインターフェース中にディスプレイされる。そのビデオのクリップからのキーフレームは、それらが発生する時間順序で示される。本発明の一実施の形態において、キーフレームのサイズは、それが表すクリップの期間に比例する。キーフレームは、スチル、パン、チルト、及びズームとしてクラス分けされたクリップに対してディスプレイされる。スチルクリップに対して、單一のキーフレーム画像は、パン、チルト及びズームに対して、画像の数が検出されたクラスタの数に依存して、ディスプレイされる。キーフレームは、このビデオが望ましくないものであると仮定して、ディフォルトによってガーベージクリップに対して示されない。にも拘らず、キーフレームインターフェースにおいて特別のセッティングは、ガーベージクリップに対してキーフレームを見ることを望むユーザに対して提供される。ステップ150において、ユーザは、ユーザインターフェースから最終出力ビデオにおいて含まれるよう記録されかビデオの部分を選択し、材

料の順序を指定することが可能となる。

【0051】図4の例示のユーザインターフェース300は、第1及び第2のインターフェースディスプレイより成る。第1のインターフェースディスプレイ310は、発生する時間順序で夫々のビデオクリップを表すキーフレームをディスプレイし、且つユーザは対応するキーフレームを選択することによって記録されたビデオからクリップを選択することが出来る。第1のインターフェースディスプレイ310中の全ての選択されたクリップは、それらのカラーヒストグラムの類似度によってクラスタリングされ、且つ類似のクリップは、複数のパイルに配置され得る。それらのパイルは、各パイル中の第1のクリップの時間順序で並んで示される。また、各パイルは、年代順である。利用可能ウインドウエリアは、ディスプレイされ得るパイルの数を決定する。

【0052】階層的クラスタリング方法は、2値クラスタツリーを生成するために使用され、これによって、ユーザが対応するレベルでクラスタツリーを介してカッティングすることによってクラスタの数と同じ程多くのファイルを選択することを可能とする。より大きなウインドウを使用することによって、より多くのファイルがディスプレイされる事が可能となり、これは真に類似する画像が共にグループ化されるというより直感的結果を導く。ファイル中の追加の画像を見るために、ユーザは、ファイル上をクリックすることによってそのファイルを拡張出来る。現在のディスプレイがフェードアウトされ、複数のファイルよりなる画像がフェードアウトされたディスプレイの中間の領域に示される。メインディスプレイをフェードアウトバックグラウンドとして示すことによって、ユーザにはドリルダウンプログラミングに対するコンテキストが提供される。拡張されたファイルのディスプレイは、複数のファイルを含んでもよい。このように、拡張されたディスプレイは、対応するフェードアウトエリアをクリックすることによって一つ又はそれより多くのレベルに変形され得る。

【0053】第2のインターフェースディスプレイ320は、出力ビデオを合成するためのストーリボードインターフェース又はタイムラインインターフェースである。第2のインターフェースディスプレイ320は、ユーザが第1のインターフェースディスプレイ310から希望のキーフレームをドラッグし、それらを第2のインターフェースディスプレイ320のタイムラインに沿って任意の順序で配置することを可能とする。キーフレームは、そのタイムラインの異なるポイントにそれらをドラッグすることによって再順序付けされ得る。一実施の形態では、第2のインターフェースディスプレイ320中のキーフレームは、水平方向空間を節約するために互い違いにされる。このように、全体のタイムラインを見るためのスクローリングがより少なくなる。他の実施の形態では、第2のインターフェースディスプレイ中のキ

ーフレームは、單一行としてディスプレイされる。

【0054】最後に、ブロック160において、知的ビデオスプライサが複数の編集ルールを適用して選択されたキーフレーム周りのビデオの適切なセクションを選択してそれらをブロックに併合して最終ステップ(ブロック)170で最終出力ビデオを生成する。即ち、代表的なキーフレームが選択され、ストーリボードインターフェースに所与の順序で配置されると、本発明は、自動的に最終出力ビデオに含まれるビデオの適切なクリップを決定する。選択されたキーフレームの各々周りに含まれるビデオの量を決定するタスクは、そのビデオの品質と隣接するキーフレームが選択されたか否かに依存する。また、これは使用される編集ルールに依存する。

【0055】一般に、ビデオ編集を経験した人々は、編集の決定を助けるために暗黙のせつの編集ルールを使用する。以下の一般的な編集ルールは、ビデオイン及びビデオアウトポイントの選択を助けるのみならず文脈的情報を使用する。(1)ショットは長過ぎず且つ短すぎない。(2)同じシーン中の2つのショット間でのカッティングを避けること。時間的制約条件を破らない限り複数のショットを併合すること。(3)過度に早いカメラ移動を有するショットを含まないこと。それらは見つめるのが困難であり、情報的ではない。(4)ショットが長すぎる場合、ショットの終り近くのサブクリップを選択すること。理由は、興味ある材料は、ショットの終わり近くに最も頻繁にある。(5)短いズーム、パン、又はチルトは、スチルで始まり且つ終る。より長いクリップは、それら自体を自己主張する。(6)ショットは、最小程度の明るさを有すべきである。

【0056】本発明によれば、以下のより特定的な編集ルールが使用される。(1)最小のショット長さは3秒であり、最大のショット長さは10秒であり、標準のショット長さは5秒である。(2)同じシーンから2つのショットが選択され、且つそれらの間のクリップが3秒未満の場合、それらのショットを併合する。全体のショットを10秒未満に維持することが必要な場合、外側端をトリムすること。(3)ガーベージとしてクラス分けされたクリップを決して含まないこと。(4)過度に長いクリップをトリムする場合、シーンの終りに向けてサブクリップを選択すること。(5)何れかの端にスチルクリップを含まない限り、5秒未満のズーム、パン又はチルトが含まれるべきではない。全長を10秒未満に維持しながら、何れかの端にスチルを含むようにクリップを延長すること。(6)ショットの平均明るさは、事前設定の閾値を越えるべきである。

【0057】制約条件満足システムがこれらの編集ルールを適用するために使用される。このアプローチは、編集ルールが素早く変化され且つ新たなルールが追加される点で有利である。更に、編集ルールは、過度の制約状況を処理するためには優先順位が付与されてよい。ま

た、このような制約条件ベースのアプローチは、ローカル及びグローバル制約条件の両方を結合出来る。例えば、個々のショットの長さは、ある長さの最終出力ビデオが希望される場合、調節され得る。最終の出力ビデオを見た後、ユーザが編集されたビデオを変更することを希望することもある。これは、ストーリボードインターフェースからキーフレームを削除し且つ新たなキーフレームを挿入することによってなされ得る。特定のクリップの期間を変更することのようなより微細な編集のために、ユーザがイン及びアウトポイントを指定出来る標準のビデオ編集インターフェースが提供されなければならない。

【0058】出力ビデオを生成するために上記アルゴリズムにおける幾つかのバリエーションが可能である。例えば、一実施の形態において、出力ビデオは、各シーンから中間セクションを選択することによって自動的に発生され得る。一実施の形態において、3秒セクションが選択され得る。他の実施の形態では、5秒セクションが選択されてもよい。ユーザの介在が許される場合、そのユーザは、各シーンから单一のキーフレームを見ることによって含まれるべきシーンを選択出来る。他の実施の形態において、出力ビデオは、セクションを、そのセクションのビデオ品質に基づいて各シーンから選択することによって自動的に発生され得る。例えば、各シーンから最も高いビデオ品質を有するクリップが選択され得る。最も高い品質ビデオクリップに基づくこの秒アプローチに適用可能な編集ルールは、上記の編集ルール（例えば、標準、最小及び最大クリップの長さが指定され、且つガーベージセクションが破棄される）に類似し、スチルカメラクリップがカメラモーションを有するクリップよりも好ましいという追加のルールを備える。上述のように、ユーザは、最終出力ビデオに含まれるべきシーンを選択出来る。

【0059】本発明の他のバリエーションにおいて、最

終出力ビデオに対してショットを選択するためのスタティックユーザインターフェースでのキーフレームディスプレイの代わりに、ダイナミックユーザインターフェースが設けられてもよい。例えば、ダイナミックユーザインターフェースを使用するユーザは、現在のクリップが最終出力ビデオからオミットされるべきか又は最終出力ビデオに含まれるべきかを指示するためにキーフレームに対応するビデオが再生されながら、一連のボタンを押すことが出来る。また、このダイナミックインターフェースは、スタティックユーザインターフェースと組合され且つそれと共に使用され得る。

【0060】本発明の好適な実施の形態の記述は、例示及び説明目的で提供された。その記述は、網羅的である事を意図するものではなく、又本発明をここで開示された正確な形態に限定する意図もない。明らかな如く、一例が上述のダイナミックユーザインターフェースであり、多くの変更例及びバリエーションが当業者にとっては明白であろう。上記の実施の形態は、本発明の原理及びその実際の用途を最もに説明するために選択されて記述され、それによって、他の当業者が種々の実施の形態に対して本発明を理解し、意図される特定の使用に適する種々の変更を可能とする。本発明の範囲は、特許請求の範囲の請求項及びそれと等価な範囲によって画定されるべきであることが意図されている。

#### 【図面の簡単な説明】

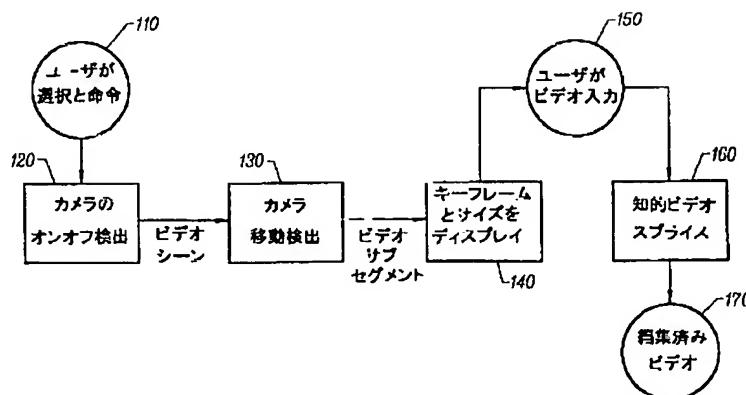
【図1】本発明のビデオ生成システムのブロック図である。

【図2】カメラモーションのタイプに基づくクラスに対応する領域に未編集ビデオをセグメント化するために使用される隠れマルコフモデル(HMM)のトップレベル構造を示す。

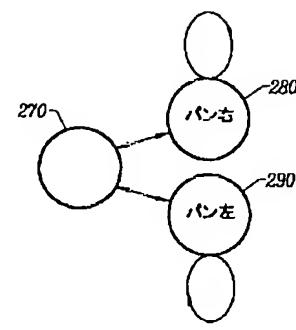
【図3】パンのサブHMMの例を示す。

【図4】第1と第2のユーザインターフェースディスプレイより成る例示のユーザインターフェースを示す。

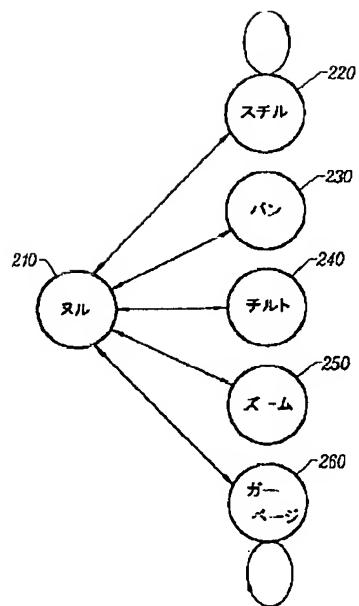
【図1】



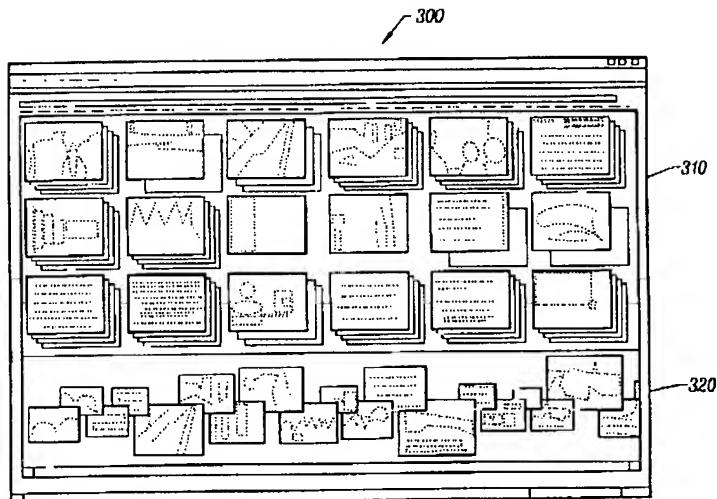
【図3】



【図2】



【図4】



---

フロントページの続き

(72)発明者 リン ディー、ウィルコックス  
アメリカ合衆国 94306 カリフォルニア  
州 パロ アルト スタンフォード アベ  
ニュー 555  
(72)発明者 ジョン エス. ボレッキー  
アメリカ合衆国 94577 カリフォルニア  
州 サン レアンドロ ダットン アベニ  
ュー 516

(72)発明者 パトリック チィーウ  
アメリカ合衆国 94025 カリフォルニア  
州 メンロ パーク ユニバーシティ ド  
ライブ 564 アパートメント 3  
(72)発明者 ジョナサン ティー、フート  
アメリカ合衆国 94025 カリフォルニア  
州 メンロ パーク ローレル ストリー  
ト 450  
Fターム(参考) 5C053 FA14 HA29